Iniziato	martedì, 7 gennaio 2020, 15:06
Stato	Completato
Terminato	martedì, 7 gennaio 2020, 15:36
Tempo impiegato	30 min. 1 secondo
Punteggio	15,00/15,00
Valutazione	**30,00** su un massimo di 30,00 (**100**%)

Domanda **1**

Risposta corretta

Punteggio ottenuto 1,00 su 1,00

How does *pruning* work when generating frequent itemsets?

Scegli un'alternativa:

- a. If an itemset is frequent, then none of its subsets can be frequent, therefore the frequencies of the subsets are not evaluated
- b. If an itemset is frequent, then none of its supersets can be frequent, therefore the frequencies of the supersets are not evaluated
- c. If an itemset is not frequent, then none of its supersets can be frequent, therefore the frequencies of the supersets are not evaluated ✔
- d. If an itemset is not frequent, then none of its subsets can be frequent, therefore the frequencies of the subsets are not evaluated

Risposta corretta.

La risposta corretta è: If an itemset is not frequent, then none of its supersets can be frequent, therefore the frequencies of the supersets are not evaluated

Domanda **2**

Risposta corretta

Punteggio ottenuto 1,00 su 1,00

Which is the main reason for the *standardization* of numeric attributes?

Scegli un'alternativa:

- a. Map all the numeric attributes to a new range such that the mean is zero and the variance is one. ✔
- b. Remove non-standard values
- c. Map all the nominal attributes to the same range, in order to prevent the values with higher frequency from having prevailing influence
- d. Change the distribution of the numeric attributes, in order to obtain gaussian distributions

Your answer is correct.

La risposta corretta è: Map all the numeric attributes to a new range such that the mean is zero and the variance is one.

Domanda **3**

Risposta corretta

Punteggio ottenuto 1,00 su 1,00

Which of the following statements regarding the discovery of association rules is true? (One or more)

Scegli una o più alternative:

- [ ] a. The support of a rule can be computed given the confidence of the rule
- [x] b. The confidence of a rule can be computed starting from the supports of itemsets ✔
- [ ] c. The confidence of an itemset is anti-monotonic with respect to the composition of the itemset
- [x] d. The support of an itemset is anti-monotonic with respect to the composition of the itemset ✔

Your answer is correct.

Le risposte corrette sono: The confidence of a rule can be computed starting from the supports of itemsets, The support of an itemset is anti-monotonic with respect to the composition of the itemset

Domanda **4**

Risposta corretta

Punteggio ottenuto 1,00 su 1,00

Given the two binary vectors below, which is their similarity according to the Jaccard Coefficient?

a b c d e f g h i j
1 0 0 0 1 0 1 1 0 1
1 0 1 1 1 0 1 0 1 0

Scegli un'alternativa:

- a. 0.5
- b. 0.2
- c. 0.1
- d. 0.375 ✔ 3/8 is the fraction of matching 1's, divided by (the number of matching 1 plus the number of non-matching)

Risposta corretta.

It is the number of matching 1 divided by the number of matching 1 + the number of non-matching

La risposta corretta è: 0.375

Domanda **5**

Risposta corretta

Punteggio ottenuto 1,00 su 1,00

What is the *cross validation*

Scegli un'alternativa:

- a. A technique to obtain a good estimation of the performance of a classifier with the training set
- b. A technique to obtain a good estimation of the performance of a classifier when it will be used with data different from the training set ✔
- c. A technique to improve the speed of a classifier
- d. A technique to improve the quality of a classifier

Risposta corretta.

La risposta corretta è: A technique to obtain a good estimation of the performance of a classifier when it will be used with data different from the training set

Domanda **6**

Risposta corretta

Punteggio ottenuto 1,00 su 1,00

Which is different from the others?

Scegli un'alternativa:

- a. Decision Tree
- b. SVM
- c. Dbscan ✔ This is not a classification method
- d. Neural Network

Risposta corretta.

La risposta corretta è: Dbscan

Domanda **7**

Risposta corretta

Punteggio ottenuto 1,00 su 1,00

Which is the main purpose of *smoothing* in Bayesian classification?

Scegli un'alternativa:

- a. Dealing with missing values
- b. Classifying an object containing attribute values which are missing from some classes in the training set ✔
- c. Reduce the variability of the data
- d. Classifying an object containing attribute values which are missing from some classes in the test set

Risposta corretta.

La risposta corretta è: Classifying an object containing attribute values which are missing from some classes in the training set

Domanda **8**

Risposta corretta

Punteggio ottenuto 1,00 su 1,00

The *information gain* is used to

Scegli un'alternativa:

- a. select the attribute which maximises, for a given test set, the ability to predict the class value
- b. select the attribute which maximises, for a given training set, the ability to predict all the other attribute values
- c. select the attribute which maximises, for a given training set, the ability to predict the class value ✔
- d. select the class with maximum probability

Your answer is correct.

La risposta corretta è: select the attribute which maximises, for a given training set, the ability to predict the class value

Domanda **9**

Risposta corretta

Punteggio ottenuto 1,00 su 1,00

Which of the following *is not* an objective of feature selection

Scegli un'alternativa:

- a. Avoid the *curse of dimensionality*
- b. Reduce time and memory complexity of the mining algorithms
- c. Select the features with higher range, which have more influence on the computations ✔
- d. Reduce the effect of noise

Risposta corretta.

La risposta corretta è: Select the features with higher range, which have more influence on the computations

Domanda **10**

Risposta corretta

Punteggio ottenuto 1,00 su 1,00

Which of the statements below is true? (One or more)

Scegli una o più alternative:

- a. K-means is very sensitive to the initial assignment of the centers ✔ No, being based on distances, if the number of attributes is very large k-means is prone to the *curse of dimensionality*
- b. Sometimes k-means stops to a configuration which does not give the minimum distortion for the chosen value of the number of clusters. ✔
- c. K-means is quite efficient even for large datasets ✔ No, k-means finds a local minimum of the distortion for an assigned number of clusters
- d. K-means always stops to a configuration which gives the minimum distortion for the chosen value of the number of clusters.

Your answer is correct.

Le risposte corrette sono: Sometimes k-means stops to a configuration which does not give the minimum distortion for the chosen value of the number of clusters., K-means is quite efficient even for large datasets, K-means is very sensitive to the initial assignment of the centers

Domanda **11**

Risposta corretta

Punteggio ottenuto 1,00 su 1,00

Given the definitions below:

- TP = True Positives
- TN = True Negatives
- FP = False Positives
- FN = False Negatives

which of the formulas below computes the *precision* of a binary classifier?

Scegli un'alternativa:

- a. TP / (TP + FN)
- b. (TP + TN) / (TP + FP + TN + FN)
- c. TP / (TP + FP) ✔ This is also called *positive predictive value*, which is the number of detected true positives divided by the total number of elements predicted as positive
- d. TN / (TN + FP)

Risposta corretta.

La risposta corretta è: TP / (TP + FP)

Domanda **12**

Risposta corretta

Punteggio ottenuto 1,00 su 1,00

Which of the following clustering methods is ***not*** based on distances between objects?

Scegli un'alternativa:

- a. DBSCAN
- b. Hierarchical Agglomerative
- c. Expectation Maximization ✔
- d. K-Means

Your answer is correct.

La risposta corretta è: Expectation Maximization

Domanda **13**

Risposta corretta

Punteggio ottenuto 1,00 su 1,00

In a dataset with D attributes, how many subsets of attributes should be considered for feature selection according to an exhaustive search?

Scegli un'alternativa:

- a. O(D)
- b. $O(2^D)$ ✔
- c. O(D!)
- d. $O(D^2)$

Risposta corretta.

La risposta corretta è: $O(2^D)$

Domanda **14**

Risposta corretta

Punteggio ottenuto 1,00 su 1,00

After fitting DBSCAN with the default parameter values the results are: 0 clusters, 100% of noise points. Which will be your next trial?

Scegli una o più alternative:

- [x] a. Reduce the minimum number of objects in the neighborhood ✔
- [ ] b. Reduce the minimum number of objects in the neighborhood and the radius of the neighborhood
- [ ] c. Decrease the radius of the neighborhood
- [x] d. Increase the radius of the neighborhood ✔

Risposta corretta.

Le risposte corrette sono: Reduce the minimum number of objects in the neighborhood, Increase the radius of the neighborhood

Domanda **15**

Risposta corretta

Punteggio ottenuto 1,00 su 1,00

In a decision tree, an attribute which is used only in nodes near the leaves...

Scegli un'alternativa:

- a. ...has a high correlation with respect to the target
- b. ...is irrelevant with respect to the target
- c. ...guarantees high increment of purity
- d. ...gives little insight with respect to the target ✔

Risposta corretta.

La risposta corretta è: ...gives little insight with respect to the target

◄ Lab Activity 17-12-2019 - Simulation of lab exa Vai a... Introduction to Big Data - Slides ►